

<https://helda.helsinki.fi>

Machine learning suggests sleep as a core factor in chronic pain

Miettinen, Teemu

2021-01

Miettinen , T , Mäntyselkä , P , Hagelberg , N , Mustola , S , Kalso , E & Lötsch , J 2021 , ' Machine learning suggests sleep as a core factor in chronic pain ' , Pain , vol. 162 , no. 1 , pp. 109-123 . <https://doi.org/10.1097/j.pain.0000000000002002>

<http://hdl.handle.net/10138/338156>

<https://doi.org/10.1097/j.pain.0000000000002002>

cc_by_nc

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Machine learning suggests sleep as a core factor in chronic pain

Teemu Miettinen¹, Pekka Mäntyselkä², Nora Hagelberg³, Seppo Mustola⁴, *Eija Kalso^{1,5}, and *Jörn Lötsch^{6,7}

¹ Pain Clinic, Department of Anesthesiology, Intensive Care, and Pain Medicine, University of Helsinki, Helsinki University Central Hospital, Helsinki, Finland

² Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland and Primary Health Care Unit, Kuopio University Hospital, Kuopio, Finland

³ Pain Clinic, Turku University Hospital, Turku, Finland

⁴ Department of Anesthesia, Intensive Care, and Pain, South Karelia Central Hospital, Lappeenranta, Finland

⁵ Sleepwell Research Programme, University of Helsinki, Helsinki, Finland

⁶ Institute of Clinical Pharmacology, Goethe - University, Theodor – Stern - Kai 7, 60590 Frankfurt am Main, Germany

⁷ Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Theodor – Stern - Kai 7, 60590 Frankfurt am Main, Germany

*Equal contribution

The manuscript contains 49 pages, including three figures and four tables.

Correspondence to: Teemu Miettinen. Address: Pain Clinic, Helsinki University Hospital, Haartmaninkatu 2A, PO Box 140, 00029 HUS Helsinki, Finland. Tel.: +358504123172. E-mail address: teemu.miettinen@helsinki.fi

Abstract

Patients with chronic pain have complex pain profiles and associated problems. Subgroup analysis can help identify key problems. We used a data-based approach to define pain phenotypes and their most relevant associated problems in 320 patients undergoing tertiary pain management. Unsupervised machine learning analysis of parameters “pain intensity”, “number of pain areas”, “pain duration”, “activity pain interference” and “affective pain interference”, implemented as emergent self-organizing maps, identified three patient phenotype clusters. Supervised analyses, implemented as different types of decision rules, identified “affective pain interference” and the “number of pain areas” as most relevant for cluster assignment. These appeared 698 and 637 times, respectively, in 1000 cross-validation runs among the most relevant characteristics in an item categorization approach in a computed ABC analysis. Cluster assignment was achieved with a median balanced accuracy of 79.9%, a sensitivity of 74.1%, and a specificity of 87.7%. In addition, among 59 demographic, pain etiology, comorbidity, lifestyle, psychological, and treatment-related variables, sleep problems appeared 638 and 439 times among the most important characteristics in 1000 cross-validation runs where patients were assigned to the two extreme pain phenotype clusters. Also important were the parameters “fear of pain”, “self-rated poor health”, and “systolic blood pressure”. Decision trees trained with this information assigned patients to the extreme pain-phenotype with an accuracy of 67%. Machine learning suggested sleep problems as key factors in the most difficult pain presentations, therefore deserving priority in the treatment of chronic pain.

Introduction

Pain intensity is a valid measure in acute pain, but its validity as the primary measure in chronic pain has been questioned [48]. A recent study found that the health-related quality of life of patients with

persistent pain is associated with multiplicity of pains, pain-related distress, and effects of pain on daily life, but not with pain intensity when other pain-related factors were taken into account [97]. Thus, the subjective health state of a pain patient is likely to depend more on the complex of pain-related factors than on the mere intensity of pain, and it is the pattern of these pain-related factors that we should address in treatment. However, the pattern of pain-related factors, or the success of treatment, may be affected by other factors as well, such as comorbidities or the psychological consequences of persistent pain.

To clarify the complex profiles of pain and associated factors, several studies have analyzed how pain patients may be grouped. A well-known line of research has identified: the “adaptive copers”, characterized by low pain intensity and pain interference, low affective distress, and high activity and life control; the “interpersonally distressed”, distinguished by high interpersonal distress; and the “dysfunctional”, characterized by high pain intensity and pain interference, high affective distress, and low activity and life control [13,94,95]. When demographics, anxiety symptoms, depression, and pain acceptance were included in the analysis, those in the group with the highest pain intensity and pain interference had also the highest number of pain areas, lowest proportion of university education, highest load of anxiety and depressive symptoms, and lowest pain acceptance [29]. Also, the lowest pain thresholds in quantitative sensory testing were found with the highest pain intensity and highest number of pain areas [22]. Taken together, a recurring finding is the identification of three or four subgroups of patients, with groups at the extreme ends with respect to pain intensity, pain interference, and vulnerability factors. However, previous research has not included variables such as comorbidities or lifestyle factors. Also, it is not clear which attributes or problems most clearly signal the probability of being at either end of the spectrum. New machine learning methods may help to clarify this.

In this study, we used a data-driven approach in a cohort of patients with persistent pain of various causes to identify the most important attributes or problems specific to different pain phenotypes. In addition to pain phenotype-related factors, 59 further variables, representing factors recognized to be, and others possibly, associated with pain phenotypes were identified, including demographic factors, pain etiology, comorbidities, lifestyle factors, psychological variables, and treatment-related factors. These data were analyzed using unsupervised and supervised machine learning methods with the aim of (i) identifying and interpreting patterns arising from different pain phenotypic parameters and (ii) selecting among the further parameters those that were informative in associating a patient with a particular pain phenotype.

Methods

Study setting and design

This was an observational cohort study in chronic pain patients treated in three multidisciplinary tertiary pain clinics in Finland. A multi-center "KROKIETA" study was used to collect patient data on a broad scale, including socioeconomic factors, lifestyle factors, psychological variables, previous treatments, and biochemical indicators [98]. The study adheres to the STROBE guidelines (www.equator-network.org) for reporting observational studies.

Participants

The cohort comprised 320 patients, enrolled between September 2013 and November 2016. Patients were invited to participate by mail with an invitation to the first appointment in the multidisciplinary pain clinics of participating hospitals (two university hospitals, one central hospital). The mailing included the usual pre-appointment questionnaires on pain, health, treatment experiences, and lifestyle factors, which the patients completed at home. The patients who agreed to participate in the study filled in additional questionnaires on psychological factors, nutrition, substance use, and quality

of life during their first visit to the pain clinic. Nurses measured patients' weight, height, and blood pressure, and physicians registered diagnoses, medications, and previous treatments. All study data were stored in an electronic database. The coordinating ethics committee of the Helsinki and Uusimaa Hospital District approved the study protocol (29/13/03/00/12). All patients agreeing to participate signed an informed consent form. The only exclusion criteria were active cancer or inability to answer questionnaires in Finnish.

Variables and measurements

The identified parameters were grouped into seven different categories comprising (i) pain phenotype-related features, (ii) pain etiology-related information, (iii) psychological parameters, (iv) demographic parameters, (v) lifestyle-related parameters, (vi) information about previous treatments, and (vii) information about comorbidities (**Table 1**).

Pain phenotype-related variables

Pain intensity, activity pain interference, and affective pain interference were assessed with the Brief Pain Inventory (BPI) [20]. The BPI quantifies pain intensity with four items (worst, least, average, right now), activity pain interference with three items (walking, work, general activity), and affective pain interference with four items (mood, relations with other people, enjoyment of life, sleep) [19]. All items are scored using an 11-point numerical rating scale ranging from 0-10. For this study, we used the mean score of the four items for pain intensity, of the three items for activity pain interference, and of the four items for affective pain interference. Several studies have validated the psychometric properties of the scales [7,19,52,88].

The number of pain areas was obtained from the pre-appointment health questionnaire, using a figure of the human body where the patient had marked all the areas with pain. For calculating the sum of pain areas, the figure was divided into 11 subareas (head, face, front of neck, back of neck, shoulders,

chest, upper limbs, stomach, upper back, lower back, and lower limbs). The duration of the pain was queried using a six-interval scale (less than a month, 1-3 months, 3-6 months, 6-12 months, 1-2 years, and more than 2 years).

Pain etiology

Pain etiology was recorded by a physician after patient examination on the first visit to the pain clinic, using the categories (i) any neuropathic pain, (ii) back pain, (iii) other musculoskeletal pain, (iv), complex regional pain syndrome (CRPS), (v) headache, (vi) phantom limb pain, (vii) fibromyalgia, (viii) chronic pain syndrome, and (ix) other pain diagnosis.

Psychological parameters

Self-rated health was queried by asking the patients to rate their current health as excellent, good, average, rather poor, or very poor [65,82]. Childhood adversities were queried as the sum of adversities present before the age of 16 years, as applied in the Health 2000 epidemiological survey [79]. Adversities included financial difficulties in the family, parental unemployment, parental serious illness, paternal alcohol abuse, maternal alcohol abuse, paternal serious psychiatric condition, maternal serious psychiatric condition, serious conflicts within family, parental divorce, own serious illness, and bullying at school.

Beck Depression Index II (BDI-II) [8] assessed depressive symptoms. BDI-II comprises 21 items, answered on a scale from 0 to 3. The psychometric properties of the BDI-II have proved reliable in patients with chronic pain [36].

Chronic Pain Acceptance Questionnaire (CPAQ) [69] measures acceptance of chronic pain. It comprises 20 items, which the patient answers on a 7-point scale ranging from “never true” to “always true”. The items divide into the subscales “activity engagement”, which refers to an outlook of keeping up with

life activities despite the presence of pain, and “pain willingness”, which reflects how much the patient considers avoiding or controlling as ineffective strategies to handle pain. Psychometric studies have found both the factor structure and the internal consistencies (Cronbach’s α [21] = 0.83 - 0.89) of the scales to be good [104].

Pain Anxiety Symptoms Scale 20 (PASS-20) was used to assess pain-related anxiety [68]. PASS-20 comprises four subscales that measure different anxiety components: cognitive anxiety, escape/avoidance behavior, fear of pain, and physiological anxiety symptoms. The measure includes 20 items, which the patient answers on a 6-point scale ranging from “never” to “always”. The factor structure has been confirmed in several samples [1,84], and internal consistencies for the scales are acceptable (Cronbach’s α = 0.75 - 0.86) [68].

Demographic parameters

Demographics were queried in the pre-appointment health questionnaire. Household income was recorded as the total income before taxes (category 1, < €10,000, followed by categories with €10,000 increments).

Lifestyle-related parameters

Body mass indices (BMI) were calculated from measurements taken by nurses during the first visit to pain clinics. Blood pressure was calculated as the average of two sedentary measurements after a 10-min rest. In addition, nurses measured waist circumference. Current smoking was assigned when the patient had smoked regularly for at least a year and continued to smoke, as queried in the nationwide health study FINRISK [77]. Exercising activity was assessed by asking patients to estimate the weekly frequency of leisure-time physical exercise producing at least a slight shortness of breath and perspiration, and lasting for at least 20 min at a time [77]. Self-reported exercise frequency has been associated with the subjective measure of fitness [38], the activity category measured by an electronic

monitor, and the maximal oxygen consumption in a treadmill test [56]. The number of daily sitting hours was queried as the sum of the patient's estimate of the time spent sitting while at work, at home watching TV or at a computer, and in a motorized vehicle during the day.

Sleep problems were assessed with five items from the Basic Nordic Sleep Questionnaire (BNSQ) [75]. Items queried difficulties in falling asleep, waking up during the night, sleep medication usage, tiredness in the morning, and tiredness during the day. Items were summed to form an index of sleep problems severity (sleep problems index). These items have been selected previously for research purposes based on clinical judgment and the criteria for insomnia in the International Classification of Diseases, Tenth Revision (ICD-10) [98].

The nutritional index was constructed to reflect adherence to the national nutritional recommendations [27] from the patient-reported consumption of fruit and vegetables, fish, whole grains, dairy products, and fats (0 indicating adherence in none of these areas and 5 indicating adherence in all). Drug abuse was assigned as none, has used, or dependency by asking "Have you ever abused drugs in your life?", and if so, "Can you manage without the drugs for at least one week?" The frequency of drinking alcohol was recorded with a question from the Alcohol Use Disorders Identification Test (AUDIT) [87] with categories "never", "once a month or less", "2-4 times a month", "2-3 times a week", or "4 times a week or more".

Previous treatments

The number of physician visits and the number of work days missed due to sickness within the previous 12 months were queried in the pre-appointment health questionnaire. The number of negative and positive treatment experiences was summed from the pre-appointment questionnaire, where the patient had indicated whether the following treatments had been used for pain, and the result (negative experience when the result had been poor or made pain worse, and positive when the result

had been mediocre to very good): guidance on pain management, nonsteroidal anti-inflammatory medication, muscle relaxant medication, opioid medication, neuropathic pain medication, physiotherapy, exercises for muscles or joints, relaxation exercises, increase in movement, hot/cold packs, pain management group, transcutaneous electrical nerve stimulation, acupuncture, voice massage therapy, chiropractor/naprapathy/osteopathy, traditional massage, trigger point manipulation, laser treatment, temporomandibular joint manipulation, occlusal splint, occlusal adjustment, cortisone injections, surgery, or treatment other than specified.

Comorbidities

Comorbidities were queried in the pre-appointment health questionnaire by asking the patient whether a physician had diagnosed or treated any of the following diseases within the previous 12 months: hypertension, heart failure, angina pectoris, diabetes, asthma, chronic obstructive pulmonary disease, rheumatoid arthritis, disease of joints other than rheumatoid arthritis, low back pain, depression, other psychiatric disorder. Additional questions included whether the patient had ever had a high level of cholesterol, was using cholesterol-lowering medications, had ever had high blood pressure, had ever used antihypertensive medication, and the type of diabetes (if diagnosed by a physician).

Data analysis

The data analysis was performed using the R software package version 3.6.1 for Linux (<http://CRAN.R-project.org/> [80]) on an Intel Core i7® - 7500U notebook computer running on Ubuntu Linux 18.04.2 64-bit.

Quantitative variables

The variables (**Table 1**) included in the data set were conceptually grouped into seven main categories. These comprised (i) $d = 5$ pain-related parameters used for the pain phenotype definition; (ii) $d = 11$ pain diagnoses such as neuropathic pain, fibromyalgia, CRPS, abdominal pain, facial pain, and others; (iii) $d = 9$ psychological factors such as those queried in the BDI or PASS inventories; (iv) $d = 8$ demographic variables such as age, gender, number of children, marital status, education, household income, and field of professional work; (v) $d = 8$ variables related to lifestyle, such as smoking, use of alcohol and illicit drugs, exercise status, and nutritional habits; (vi) $d = 4$ variables related to previous treatments, such as the number of therapies having made pain worse, or the frequency of seeing a doctor; and (vii) $d = 14$ possible comorbidities such as diabetes, hypertension, heart failure, or asthma. Variables with more than 20% missing information were excluded from the analyses.

Data analysis strategy

The data analysis was performed in two main steps (**Figure 1**). The **first step** focused on the pain phenotype and included (i) the unsupervised evaluation of a pain phenotype cluster structure emerging in the $d = 5$ pain-related parameters described above. Following the identification of the pain-related clusters in the high-dimensional data set, (ii) an interpretation of the cluster structure was sought. Therefore, supervised algorithms were applied to allow a correct summary interpretation of the pain-related phenotypes. The procedure was considered successful if (i) the phenotypes were captured by the interpretation with high (balanced) accuracy and (ii) the phenotypes reflected a relevant clinical framework. Therefore, algorithms of explainable artificial intelligence (XAI) [6] were used as this type of algorithm creates sets of simple rules that mimic the clinical decision process and make the class assignment and the relevant items underlying the clusters transparent [32] to the pain experts.

The **second step** of the data analysis was aimed at identifying associations of the other parameters with the identified pain-phenotypes, i.e., it aimed at a mapping of parameters from the six other categories, not directly derived from pain, to the pain phenotype clusters. Therefore, supervised methods were used.

The analyses also included selection of parameters that were relevant in the context of the pain phenotype clusters, both directly among the five pain-related parameters, and indirectly among the further parameters. Therefore, feature selection steps were included [86], where parameters that did not provide useful information for phenotype cluster assignment were removed from subsequent analyses.

Identification of pain phenotype cluster structure

Cluster analysis of pain-related parameters

The pain phenotype was searched in a cluster structure in the data space given by the $d = 5$ pain-related features. The correlation structure among the features was assessed by calculating Spearman's ρ [91]. In the case of relevant correlations, uncorrelated variables were obtained by orthogonal projection of the data onto a lower-dimensional linear space, i.e. the principal subspace, so that the variance of the projected data was maximized [42,76]. The PCA was performed using unit variance scaling including centering of the data. Of the resulting principal components (PCs), those with eigenvalues > 1 were retained [35,47]. Subsequent cluster analysis was performed according to the workflow proposed elsewhere [50], i.e. cluster analysis was carried out on the PCs and implemented as hierarchical clustering using Ward's method [102] and the Euclidean distance. The clusters were subsequently consolidated using k-means clustering [63] to improve the initial partition obtained from hierarchical clustering [50]. As default in the R library, a maximum of 10 iterations was allowed for the latter procedure. These calculations were performed using the R libraries "FactoMineR"

(<https://cran.r-project.org/package=FactoMineR> [58] and “cluster” (<https://cran.r-project.org/package=cluster> [64])).

Verification of the number of clusters, assessment of the stability of the cluster allocation of the patients, and assessment of the cluster quality were performed using 1000 times repeated cluster analyses of random data sets obtained by means of Bootstrap resampling [25] from the original data set. The most consistent cluster number was determined by automatic recognition implemented in the library “FactoMineR”. In addition, a total of 30 different indices available for this purpose was calculated using the R library “NbClust” (<https://cran.r-project.org/package=NbClust> [18]). The final number of clusters, among [2,...,15] clusters allowed, was determined by following the majority rule, i.e. (i) the cluster count proposed by the largest number of indices and (ii) the most common number of clusters obtained during the 1000 runs. Cluster stability was evaluated by calculating the adjusted Rand index [81] between the original cluster assignment of the resampled data and a new cluster assignment obtained with the same clustering strategy as described above. Cluster quality was evaluated by calculating the Silhouette index [85]. The calculations were performed using the R libraries “cluster” and “fossil” (<https://cran.r-project.org/package=fossil> [99]). The 95% confidence intervals of the estimated cluster stability and quality criteria were obtained as the 2.5th and 97.5th percentiles of the values obtained in the 1000 runs.

Interpretation of pain phenotype cluster structures

After evaluations of cluster number, stability, and quality, the clusters were interpreted based on the original variables used for cluster analysis. Using supervised machine learning techniques [62], the relevant variables underlying the resulting cluster structure were selected, using the raw pain-related variables to create rules suitable for a comprehensible cluster interpretation. That is, the aim of this procedure was to reduce the data space to relevant components and to create simple and

understandable rules for the assignment of a patient to the correct phenotype cluster. Ideally, only a few short rules should be created.

Therefore, a feature selection step was performed using filter techniques [86] based on an item categorization technique into relevant and less important features. This was implemented as computed ABC analysis [96], which aims to divide a set of positive numerical data into three disjoint subsets called “A”, “B”, and “C”. Set “A” should contain the “important few”, i.e. those elements that make it possible to obtain a maximum yield with a minimum effort [46,74]. These calculations were done using our R package “ABCanalysis” (<http://cran.r-project.org/package=ABCanalysis> [96]). The numerical data submitted to computed ABC analysis consisted of a measure of feature importance, obtained as the decrease in classification accuracy when the feature had been omitted from classifier building. This was repeated 1000 times on data randomly resampled from the original data set (Monte Carlo resampling [34]) into disjoint training (2/3 of the data) and test (1/3 of the data) data subsets. The final size of the feature set corresponded to the most common size of set “A” in the 1000 runs. The members of the feature set were selected in descending order of their appearances in ABC set “A”.

With an emphasis on cluster interpretability, the classifiers were deliberately chosen to be symbolic [73], i.e. the decision on how to obtain a classification can be interpreted by a domain expert from a combination of conditions on the features. This is in line with current informatics research efforts to make machine-learned algorithms explainable [71]. Typical implementations are hierarchical “if-then-else” rules assembled in decision trees, which are currently used (i) in the form of classification and regression trees (CART [12]) and (ii) as rule sets based on partial decision trees (PART), which focus on simple and short rules in the form of PART decision lists [28]. These calculations were carried out using the R packages “rpart” (<https://cran.r-project.org/package=rpart> [93]) and “RWeka” (<https://cran.r-project.org/package=RWeka> [40]). Hyperparameter tuning had indicated the use of Gini impurity [33]

as the split criterion for CART. In addition, the requirement was that a minimum number of 20 observations had to be in a node to try a split, and a maximum depth of the tree of 30 successive splits was allowed. In PART, the rules were pruned with a confidence threshold of 0.25, being the default setting in the used library.

However, since simple understandable rules can lead to simplified and thus poorly functioning classifiers, their performance was tested against a standard subsymbolic classifier, i.e. where a potentially better performance is sought by forgoing the possibility of easily obtaining biomedical explanations for the functioning of the algorithm [62]. Random forests [11,39] are qualified for this task because they provide a commonly used standard classifier that typically uses hundreds of simple trees. The class assignment is obtained by a majority vote that conceals the decision-making process from immediate topical interpretation. Hyperparameter tuning assessments had indicated the use of \sqrt{d} parameters for each tree in the random forest, which corresponds to the standard procedure implemented in the R library “randomForest” (<https://cran.r-project.org/package=randomForest> [61]). Using this setting, the out-of-bag error decreased up to forest sizes of approximately 300 - 500; however, forest sizes of 1500 trees were used in further analyses since this has been shown to not be penalized with overfitting [92]. Furthermore, to also include a more common classification procedure known from statistical data analyses, a regression analysis was performed. This was implemented as multinomial log-linear regression using the R library “nnet” that fits the model via neural networks (<https://cran.r-project.org/package=nnet> [100]).

The performance of the obtained classifiers to assign patients to the correct phenotype clusters was evaluated by running the classifiers on 1000 different disjoint training and test data sets generated by Monte Carlo resampling [34] to split the original data set into a training (2/3 of the data) and a disjoint test (1/3 of the data) data subset. To further address possible overfitting, a negative control condition was created by randomly permuting each feature in the training data subset. It was expected that a

classifier trained with these data will merely provide a 50% balanced classification accuracy, equivalent to a coin toss. Classification performance was judged mainly by balanced accuracy [14]. In addition, further standard measures of classification performance were calculated, such as sensitivity, specificity, positive and negative predictive values, precision, recall [4,5], and the F1 score [44,90]. The performance measures were calculated using the R library “caret” (<https://cran.r-project.org/package=caret> [55]). In addition, the area under the receiver operating characteristic (ROC) curve, AUC-ROC, was calculated using the “multiclass.roc” procedure implemented in the R library “pROC” (<https://cran.r-project.org/package=pROC> [83]). Further measures of classifier performance were implemented as discriminant power [9,89] and Youden’s index [107]. Finally, standard statistical comparisons of parameter values among clusters were performed using Kruskal-Wallis tests [54]. The α level was set at 0.05 and corrected for multiple testing according to Bonferroni’s proposal [10].

Mapping of other phenotypic factors on the pain phenotype cluster structure

A mapping of the six other categories of phenotypic parameters on the pain phenotype clusters was performed using the same supervised methods as described above. Categorical variables with more than one category were “one-hot” encoded, i.e. for each category a new [0,1] scaled variable, indicating “yes” or “no”, respectively, was created. Missing values (**Table 1**) in numeric or nominal data were replaced by the median or mode, respectively, of the parameter values, but only for those variables with $\leq 20\%$ missing values; otherwise, the variable was omitted from the analysis. Considering the heterogeneity of parameters (including interval scaled, ordinal scaled, categorical, and binary variables), tree-based classifiers were again used, which also provided analytical consistency with the pain phenotype interpretation assessments. Similarly, the focus was again on XAI aimed at providing the class assignment which would be immediately transparent and reproducible for a pain expert.

Results

The analyzed cohort consisted of $n = 277$ patients (age: range 18 – 77 years, mean \pm standard deviation: 45.7 ± 13.2 years, sex: 96 men, 181 women) for whom key pain, demographic, and psychology-related data were complete (**Figure 1** and **Table 1**). Variables concerning the use of blood pressure medication, nutritional index, drug abuse, and alcohol consumption frequency had $> 20\%$ missing values and were therefore omitted from the analyses.

Pain phenotype cluster structure

Number and cluster stability of pain phenotype clusters

Among the $d = 5$ pain phenotype-related parameters (number of pain areas, duration of pain, pain intensity, affective pain interference, and activity pain interference), the first two parameters, i.e. the number of pain areas and the duration of pain, were weakly correlated (Spearman's $\rho = 0.2$), while the last three, i.e., pain intensity, affective pain interference and interference with activities, showed moderate or higher correlations (pain intensity and affective pain interference: $\rho = 0.58$, pain intensity and activity pain interference: $\rho = 0.59$, affective pain interference and activity pain interference: $\rho = 0.65$). In addition, the number of pain areas was weakly but significantly correlated with pain intensity and affective pain interference ($\rho = 0.11$ and 0.16 , respectively). Therefore, principal component analysis (PCA) was performed, which resulted in two PCs with eigenvalues > 1 (**Table 2**) that explained 45.9% and 22.9% of the total variance in the pain-related parameters, respectively. PC1 carried loadings mainly from the parameters derived from BPI, while PC2 carried mainly loadings from the number of pain areas and pain duration. It is noteworthy that the correlation analyses were carried out on the pain-phenotype -related data with a view to the subsequent cluster analysis in order to avoid redundant components and not to analyze a general correlation structure of the whole data set.

The latter would have provided only limited information given the very heterogeneous, non-pain-related data, including one-hot encoded nominal data (**Table 1**).

Hierarchical Ward clustering with subsequent consolidation using k-means clustering identified $k = 3$ clusters (

Figure 2 A) as the most common solution among 1000 retries with randomly resampled data. In particular, $k = 3, \dots, 8$ clusters had resulted in 56.6, 37.2, 0.44, 0.11, 0.04, and 0.03% of the runs according to the cluster number by the “HCPC” procedure implemented in the R library “FactoMineR”. A number of $k = 3$ clusters was also the majority vote of the 30 different indices to determine the number of clusters. Clusters #1, #2, and #3 included 81, 109, and 87 subjects, respectively. The use of $k = 3$ clusters in 1000 runs on random Bootstrap resampled data yielded a median adjusted Rand index of 0.49 (95% confidence interval, CI: 0.17 - 0.89) and a median Silhouette index of 0.35 (95% CI: 0.3 - 0.41).

Interpretation of pain phenotype clusters

Feature selection based on the computed ABC analysis of the decrease in classification accuracy when the feature was omitted resulted for all three algorithms, i.e. CART, PART and RF, in $d = 2$ pain-related parameters as the most frequent size of set “A”. That is, its most frequent members were “affective pain interference” and the “number of pain areas”. For CART, which was later chosen as the basis for cluster interpretation, the “affective pain interference” was part of set “A” 693 times in the 1000 repeated runs with data subsets taken at random from the original data set, while the “number of pain areas” was in set “A” 637 times.

Trees or rule sets built from these two parameters were able to assign patients to their correct phenotype at a median balanced accuracy, across the three clusters, of 79.9% and 80.4% for CART and PART, respectively (**Table 3**). This was only slightly outperformed by RF (81.3% balanced accuracy), or

by multinomial regression (82.1% balanced accuracy). In contrast, training the algorithms with permuted data resulted in balanced accuracies of approximately 50%. However, the reduction to two features was penalized by a reduction in classification accuracy. Classifiers built with the full set of $d = 5$ pain phenotype-related parameters provided approximately 10% more median balanced classification accuracy (**Table 3**). Consistently, all pain phenotype-related parameters differed significantly between the three clusters (

Figure 2 C). Kruskal-Wallis tests were significant for the number of pain areas ($\chi^2 = 87.314$, $df = 2$, $p < 2.2 \cdot 10^{-16}$), duration of pain ($\chi^2 = 60.258$, $p = 8.227 \cdot 10^{-14}$), pain intensity ($\chi^2 = 111.86$, $p < 2.2 \cdot 10^{-16}$), affective pain interference ($\chi^2 = 157.11$, $p < 2.2 \cdot 10^{-16}$), and activity pain interference ($\chi^2 = 132.52$, $p < 2.2 \cdot 10^{-16}$), i.e. all p-values were below the corrected α level of 0.01.

CART provided a total of seven rules for the assignment of patients to the three pain phenotype clusters (

Figure 2 C), whereas PART provided nine rules. The RF algorithm was not designed to provide simple interpretations of the classification process. As CART was only narrowly outperformed by PART in the above performance analyses, the desired simplicity in the class assignment rules was considered as best obtained using the reduced feature set and the decision tree-based rules obtained with the CART algorithm. For example, the CART-based rules (

Figure 2 B) assigned a subject either to pain phenotype cluster #1 or to cluster #3, depending on the value of affective pain interference (< 6.4 or ≥ 6.4). Whether the subject was finally moved to cluster #2 depended on combinations of the two parameters selected for the final rule set. Class assignment was not perfect as indicated by the number of correctly or incorrectly classified patients shown in the bottom leaves of the decision tree in

Figure 2 B and also not equally good for each of the three clusters. Specifically, class assignment succeeded at median balanced accuracies of 86.6%, 70.9%, and 83.7% for clusters #1, #2, and #3, respectively (upper part of **Table 4**).

Signal parameters for core factors in chronic pain phenotypes

In line with the better performance of the assignment to pain phenotype clusters #1 (lowest pain intensity and pain interference) and #3 (high pain intensity and pain interference, and the greatest number of pain areas) through the use of pain phenotype-related parameters, the same clusters were also more accessible through hierarchical rules built from other categories of phenotypic parameters (**Table 4**). Moreover, for the three-cluster solution, or for cluster #2 versus the two other clusters, the confidence interval of the balanced accuracy occasionally included 50%, so it could not be clearly established that some classifiers exceeded guesswork in this task. The analysis therefore focused on the pain phenotypes #1 and #3, both times against the two other phenotypes clusters combined. Both PART and RF required 11 to 15 parameters for only slightly better accuracy than CART, which required $d = 3$ parameters (**Table 4**).

Therefore, CART-based rules were again considered best suited to meet the desired simplicity in cluster interpretation. The tested scenarios provided the phenotypic parameters, not directly related to pain, that were most informative for the assignment of a patient to the extreme phenotypes in the present cohort, i.e. pain-based clusters #1 (lowest pain intensity and pain interference) or #3 (high pain intensity and pain interference, and the greatest number of pain areas) (**Figure 3 A and B**, respectively). Among them, the index “sleep problems” and “PASS fear of pain” appeared most frequently in the decision trees. That is, among the 1000 runs performed with CART to map feature space on cluster #1 versus the other clusters, “sleep problems” was member of the ABC set “A” 638 times and “PASS fear of pain” 406 times. For cluster #3 versus the other clusters, “sleep problems” was member of set “A” 439 times, and “PASS fear of pain” 401 times. In both scenarios, three characteristics were the most

frequently observed size of set "A" (**Table 4**). For assignment to cluster #1, systolic blood pressure played an additional role, whereas for assignment to cluster #3, self-rated health was important.

The CART-based rules (**Figure 3**) assigned a subject to pain phenotype #3, depending on the value of "sleep problems" (≥ 15) and on either a rating of "fear of pain" ≥ 15 or a rating of "self-rated health" ≥ 4.5 . However, the class assignment was weak, as indicated by the number of correctly or incorrectly classified patients shown in the bottom leaves of the decision trees, and moreover, by the modest values of classification accuracy which did not exceed a median of 65% (lower part of **Table 4**).

Discussion

The analysis revealed three clusters, with two clusters in the extremes with regard to pain-related factors, a finding replicated in previous research [22,29,94]. Cluster #1 had the lowest pain intensity and pain interference, while cluster #3 showed a combination of high pain intensity and high pain interference, and the greatest number of pain areas. Among the pain-related factors, the variables with the highest accuracy in predicting group membership were affective pain interference and the number of pain areas. Among the rest of the variables, the sleep problems factor was the most important for assigning a patient to the extreme phenotypes, followed by fear of pain, self-rated health, and systolic blood pressure.

Affective pain interference over pain intensity

Interestingly, affective pain interference (pain disrupting enjoyment of life, social relationships, mood, and sleep) emerged as the most distinctive factor for the pain phenotypes, while pain intensity did not. Research has associated more difficult pain with decreased life satisfaction [16], increased negative emotions [30], and increased sleep problems [43,101]. In addition, brain imaging findings suggest that, as pain becomes chronic, brain activity shifts from sensory regions involved in acute pain to emotion-related regions [37]. Thus, pain experience in chronic pain may be associated more with the overall

negative sequelae that impact emotions than with the mere intensity of pain. Affective pain interference probably captures the continuum from least to most difficult pain better than pain intensity ratings.

Common underlying processes?

Greater affective pain interference, more pain areas, more sleep problems, greater fear of pain, and lower self-rated health emerged as the distinctive combination of features for patients in cluster #3 with the most severe pain presentation (high pain intensity and pain interference, and the greatest number of pain areas). These features among all parameters recorded could be used to assign patients to the identified pain clusters. However, all classifiers that were trained to make these assignments were complex. Thus, no single feature could be separated out alone for cluster assignment: each was always found in combination in the different rule sets used in the different classifiers. Nevertheless, hierarchical classifiers, such as the CART classifier, allow the features to be ranked in the top-down direction of the decision tree, suggesting that sleep problems were the most important among the parameters not directly related to pain, since the initial assignment decision was always made on this basis, while further refinements of cluster assignment were made on the basis of other characteristics. The scope of this study was not to investigate the relationships between these factors, but the question emerges of whether these factors are indicative of key processes that are closely linked and that exacerbate persistent pain. More sleep problems, greater fear of pain, and lower self-rated health have been associated with recurring, more intense, more interfering, and more widespread pain in previous research as well [22,29,43,65,82,101,105].

Evolutionarily, pain is a strong signal to prepare for a threat with both physiological (arousal) and cognitive (attentional demand) reactions [24]. Anxiety, such as fear of pain, may lead to further arousal, setting the stage for problems like prolonged stress or sleeping difficulties. Polysomnographic recordings have revealed that anxiety increases both the latency of sleep onset and the time spent in

lighter sleep stages [41]. Continuing problems with sleep may predispose to more pain. In a diary study, worse sleep quality increased the following day's pain intensity, pain interference, negative affect, and need for rest [31]. Problems with sleep may also lead to more anxiety. Sleep disturbance has predicted the emergence of generalized anxiety, once exposed to a stressful life period [49]. Thus, pain, anxiety, and sleep problems may amplify each other, creating a vicious circle for the pain patient.

A greater number of pain areas may indicate the involvement of central sensitization, i.e. nociceptive neurons increasing their responsiveness and endogenous pain control mechanisms working less effectively. Recent research has also suggested that pain may spread because of neuroinflammatory processes [45]. In both processes, sleep problems may play a role [3,26,66].

The finding that sleep problems and fear of pain appear as distinctive features for cluster #1 as well (with least pain and pain interference) seems to provide more evidence for their importance. Perhaps their absence serves as a protective factor even if pain persists. Fear of pain has held a strong foothold in chronic pain research for years, and its associations with pain intensity, pain interference, and disability have appeared frequently [59]. Sleep problems have predicted the development of chronic pain [57]. In chronic pain, better sleep has predicted lessening of pain and pain-related anxiety [23,60]. Diagnosed sleep disorders have appeared among 44% of patients with chronic pain, with problems occurring in sleep continuity (delay in falling asleep, less time spent asleep, nightly awakenings) more than in sleep architecture [67]. Research on pain and sleep is moving its focus from mere associations to the underlying mechanisms [26]. This research may gain new momentum from discoveries such as the glymphatic system and its function in the clearance of cellular waste products from the brain [72], with sleep possibly being an important facilitator in the process [106].

Among the distinctive features in cluster #1, with least pain and pain interference, was the absence of high systolic blood pressure. Chronic pain patients are at an increased risk of high blood pressure, with evidence of reduced baroreflex sensitivity contributing to the risk of hypertension and predisposing to

more pain [15]. Normal activation of the baroreceptors initiates descending pain inhibitory pathways, thus producing analgesia, which helps to restore arousal levels in acute pain. Preserved normal regulation in the cardiovascular system may then protect pain patients from worse pain.

Surprising no-show: depression

It may seem surprising that some factors with high prevalence among those with difficult pain, like depression, did not appear in the list of the most important factors in the analysis. Bearing in mind that the analysis did not investigate cluster differences within the factors, but rather the most relevant factors for assigning a patient to a pain phenotype, this merely indicates that the selected factors were better in this task.

In the case of depression, one explanation may be that sleep problems and fear of pain may predispose a subgroup of pain patients to more severe pain, a portion of which then also develop depressive symptomatology. Research has suggested that insomnia is a stronger predictor of depression than vice versa [17]. Also, anxiety problems may more often precede the development of persistent pain than does depression [53].

Strengths and limitations

The main strength of this study is the broad inclusion of variables, reflecting many different lines of research on vulnerability to chronic pain. Since chronic pain has proven to be complex, it is likely that highly multifactorial data will be necessary in future research. Consequently, methods like the machine learning used here will be needed to select the most informative of these factors. In this study, the analysis was able to provide information on the most relevant factors for pain phenotypes. However, when using non-pain phenotype variables, the final algorithm performed only modestly. This may be related to the choice of a simple algorithm; however, using the random forest algorithm did not convincingly improve performance. Hence, more factors than those included in the present data set

seem necessary to increase algorithm performance. Finally, while the study cohort was reasonably large, there were some variables that only a few respondents indicated as positive (e.g. occupation in agriculture), and machine learning could not make use of these variables.

The pain-related data showed a considerable amount of “noise” (**Figure 2 A and C**), which potentially challenges the clustering of this data set. In fact, the question of whether cluster structures are present in a data set or not (cluster capability) is not easy to answer. For an up-to-date overview of the methods used to decide on clustering capability, see [2]. According to the criteria suggested there, the data set with the five pain-related parameters is clearly clusterable. First, the use of a suitable projection (PCA) structure indicated a cluster structure (**Figure 2 A**). Secondly, the various measurements of optimal cluster numbers indicated clusters with a considerable agreement in pointing to three clusters determined in the 1000 runs on randomly selected data subsets. Third, the statistical tests for differences between clusters were highly significant for all pain-related parameters. Fourth, a synoptic view of the n-dimensional data clearly shows clustering (**Figure 2 C**). Moreover, in addition to the criteria list mentioned above, the consistent selection of only two of the five pain-related parameters, in their original raw version and not in their PCA projection, as informative for the cluster structure further contradicts the view that all data were only unstructured noise.

The combination of pain clusters for their tree-based interpretation by parameters not directly related to pain could also be criticized. As reported in the results section, the confidence interval of the balanced accuracy of the CART classifier was occasionally 50% for the three-cluster solution. However, analyses with trained 3-class CART showed that sleep problems were again at the top of the tree, thus being the most important parameter for cluster assignment (Supplementary Figure 1). Hence, the main conclusion of their particular importance remains valid. Therefore, a re-analysis according to a previously proposed approach such as the so-called “Set Covering Machine” [51] was not further considered.

Conclusions

With the increasing number of recognized factors involved in the clinical picture of persistent pain, selecting the most informative ones becomes important. This data-driven analysis shows one approach to this problem with the use of machine-learning methodology. The results reveal that, among the several psychological, lifestyle, and other variables included in the study, sleep problems best capture the association with the extreme phenotypes of pain. Sleep problems thus warrant priority in the treatment of chronic pain.

Acknowledgments

We thank registered nurse Minna Kymäläinen and the physicians and nurses of the respective pain clinics for their contributions to this study.

Funding

The work was supported by the Governmental Research Funding (TYH2014214) (EK), the Signe & Ane Gyllenberg Foundation (EK), and Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), Zentrum: Translational Medicine and Pharmacology (JL). The funders had no role in method design, data selection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest statement

Dr. Kalso reports personal fees from Orion Pharma, personal fees from Pfizer, outside the submitted work; all other authors have no conflicts of interest to declare.

References

- [1] Abrams MP, Carleton RN, Asmundson GJ. An exploration of the psychometric properties of the PASS-20 with a nonclinical sample. *J Pain* 2007;8:879-886.
- [2] Adolfsson A, Ackerman M, Brownstein NC. To cluster, or not to cluster: An analysis of clusterability methods. *Pattern Recognition* 2019;88:13-26.
- [3] Albrecht DS, Forsberg A, Sandstrom A, Bergan C, Kadetoff D, Protsenko E, Lampa J, Lee YC, Hoglund CO, Catana C, Cervenka S, Akeju O, Lekander M, Cohen G, Halldin C, Taylor N, Kim M, Hooker JM, Edwards RR, Napadow V, Kosek E, Loggia ML. Brain glial activation in fibromyalgia - A multi-site positron emission tomography investigation. *Brain Behav Immun* 2019;75:72-83.
- [4] Altman DG, Bland JM. Diagnostic tests 2: Predictive values. *BMJ* 1994;309:102.
- [5] Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ* 1994;308:1552.
- [6] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, Garcia S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 2020;58:82-115.
- [7] Atkinson TM, Rosenfeld BD, Sit L, Mendoza TR, Fruscione M, Lavene D, Shaw M, Li Y, Hay J, Cleeland CS, Scher HI, Breitbart WS, Basch E. Using confirmatory factor analysis to evaluate construct validity of the Brief Pain Inventory (BPI). *J Pain Symptom Manage* 2011;41:558-565.
- [8] Beck AT, Steer RA, Ball R, Ranieri W. Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients. *J Pers Assess* 1996;67:588-597.
- [9] Blakeley DD, Oddone EZ, Hasselblad V, Simel DL, Matchar DB. Noninvasive carotid artery testing. A meta-analytic review. *Ann Intern Med* 1995;122:360-367.

- [10] Bonferroni C. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936;8:3-62.
- [11] Breiman L. Random Forests. Mach Learn 2001;45:5-32.
- [12] Breiman L, Friedman J, Stone C, Olshen R. Classification and Regression Trees. Boca Raton: Chapman and Hall, 1993.
- [13] Broderick JE, Junghaenel DU, Turk DC. Stability of patient adaptation classifications on the multidimensional pain inventory. Pain 2004;109:94-102.
- [14] Brodersen K, Ong C, Stephan K, Buhmann J. The balanced accuracy and its posterior distribution. Proceedings of the 2010 20th International Conference on Pattern Recognition; 2010: 3121-3124.
- [15] Bruehl S, Olsen RB, Tronstad C, Sevre K, Burns JW, Schirmer H, Nielsen CS, Stubhaug A, Rosseland LA. Chronic pain-related changes in cardiovascular regulation and impact on comorbid hypertension in a general population: the Tromso study. Pain 2018;159:119-127.
- [16] Budh CN, Osteraker AL. Life satisfaction in individuals with a spinal cord injury and pain. Clin Rehabil 2007;21:89-96.
- [17] Buysse DJ, Angst J, Gamma A, Ajdacic V, Eich D, Rossler W. Prevalence, course, and comorbidity of insomnia and depression in young adults. Sleep 2008;31:473-480.
- [18] Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. Journal of Statistical Software 2014;61:1-36.
- [19] Cleeland CS, Nakamura Y, Mendoza TR, Edwards KR, Douglas J, Serlin RC. Dimensions of the impact of cancer pain in a four country sample: new information from multidimensional scaling. Pain 1996;67:267-273.

- [20] Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore* 1994;23:129-138.
- [21] Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
- [22] Cruz-Almeida Y, King CD, Goodin BR, Sibille KT, Glover TL, Riley JL, Sotolongo A, Herbert MS, Schmidt J, Fessler BJ, Redden DT, Staud R, Bradley LA, Fillingim RB. Psychological profiles and pain characteristics of older adults with knee osteoarthritis. *Arthritis Care Res (Hoboken)* 2013;65:1786-1794.
- [23] Davies KA, Macfarlane GJ, Nicholl BI, Dickens C, Morriss R, Ray D, McBeth J. Restorative sleep predicts the resolution of chronic widespread pain: results from the EPiFUND study. *Rheumatology (Oxford)* 2008;47:1809-1813.
- [24] Eccleston C, Crombez G. Pain demands attention: a cognitive-affective model of the interruptive function of pain. *Psychol Bull* 1999;125:356-366.
- [25] Efron B, Tibshirani R. *An Introduction to the Bootstrap*. San Francisco: Chapman and Hall, 1995.
- [26] Finan PH, Goodin BR, Smith MT. The association of sleep and pain: an update and a path forward. *J Pain* 2013;14:1539-1552.
- [27] Fogelholm M, Hakala P, Kara R, Kiuru S, Kurppa S, Kuusipalo H, Laitinen J, Marniemi A, Misikangas M, Roos E, Sarlio-Lähteenkorva S, Schwab U, Virtanen S. *Terveyttä ruoasta! Suomalaiset ravitsemussuosituks* 2014. Helsinki, Finland: Valtion ravitsemusneuvottelukunta, 2014.
- [28] Frank E, Witten I. Generating Accurate Rule Sets Without Global Optimization. *Proceedings of the ICML*; 1998; 144-151.

- [29] Gerdle B, Akerblom S, Stalnacke BM, Brodda Jansen G, Enthoven P, Ernberg M, Dong HJ, Ang BO, Boersma K. The importance of emotional distress, cognitive behavioural factors and pain for life impact at baseline and for outcomes after rehabilitation - a SQRP study of more than 20,000 chronic pain patients. *Scand J Pain* 2019;19:639-711.
- [30] Gerhart JJ, Burns JW, Bruehl S, Smith DA, Post KM, Porter LS, Schuster E, Buvanendran A, Fras AM, Keefe FJ. Variability in negative emotions among individuals with chronic low back pain: relationships with pain and function. *Pain* 2018;159:342-350.
- [31] Gerhart JJ, Burns JW, Post KM, Smith DA, Porter LS, Burgess HJ, Schuster E, Buvanendran A, Fras AM, Keefe FJ. Relationships Between Sleep Quality and Pain-Related Factors for People with Chronic Low Back Pain: Tests of Reciprocal and Time of Day Effects. *Ann Behav Med* 2017;51:365-375.
- [32] Gigerenzer G, Todd PM. Fast and frugal heuristics: The adaptive toolbox. In: Gigerenzer G, Todd PM, ABC Research group. Simple heuristics that make us smart. New York, NY, United States: Oxford University Press; 2000:3-34.
- [33] Gini C. Variabilità e Mutabilità. Bologna, Italy: Tipografia di Paolo Cuppini, 1912.
- [34] Good P. Resampling methods. Boston: Birkhäuser, 2006.
- [35] Guttman L. Some necessary conditions for common-factor analysis. *Psychometrika* 1954;19:149-161.
- [36] Harris CA, D'Eon JL. Psychometric properties of the Beck Depression Inventory--second edition (BDI-II) in individuals with chronic pain. *Pain* 2008;137:609-622.

[37] Hashmi JA, Baliki MN, Huang L, Baria AT, Torbey S, Hermann KM, Schnitzer TJ, Apkarian AV.

Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. *Brain* 2013;136:2751-2768.

[38] Hassmen P, Koivula N, Uutela A. Physical exercise and psychological well-being: a population study in Finland. *Prev Med* 2000;30:17-25.

[39] Ho TK. Random Decision Forests. *Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)*: IEEE Computer Society; 1995:278.

[40] Hornik K, Buchta C, Zeileis A. Open-source machine learning: R meets Weka. *Comput Stat* 2009;24:225-232.

[41] Horvath A, Montana X, Lanquart JP, Hubain P, Szucs A, Linkowski P, Loas G. Effects of state and trait anxiety on sleep structure: A polysomnographic study in 1083 subjects. *Psychiatry Res* 2016;244:279-283.

[42] Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 1933;24:417-441.

[43] Jamison RN, Rudy TE, Penzien DB, Mosley TH. Cognitive-behavioral classifications of chronic pain: replication and extension of empirically derived patient profiles. *Pain* 1994;57:277-292.

[44] Jardine N, van Rijsbergen C. The use of hierarchic clustering in information retrieval. *Information Storage and Retrieval* 1971;7:217-240.

[45] Ji RR, Nackley A, Huh Y, Terrando N, Maixner W. Neuroinflammation and Central Sensitization in Chronic and Widespread Pain. *Anesthesiology* 2018;129:343-366.

[46] Juran J. The Non-Pareto Principle; Mea Culpa. *Quality Progress* 1975;8:8-9.

- [47] Kaiser H. The varimax criterion for analytic rotation in factor analysis. *Psychometrika* 1958;23:187-200.
- [48] Kaiser U, Kopkow C, Deckert S, Neustadt K, Jacobi L, Cameron P, De Angelis V, Apfelbacher C, Arnold B, Birch J, Bjarnegard A, Christiansen S, C de C Williams, A, Gossrau G, Heinks A, Huppe M, Kiers H, Kleinert U, Martelletti P, McCracken L, de Meij N, Nagel B, Nijs J, Norda H, Singh JA, Spengler E, Terwee CB, Tugwell P, Vlaeyen JWS, Wandrey H, Neugebauer E, Sabatowski R, Schmitt J. Developing a core outcome domain set to assessing effectiveness of interdisciplinary multimodal pain therapy: the VAPAIN consensus statement on core outcome domains. *Pain* 2018;159:673-683.
- [49] Kalmbach DA, Abelson JL, Arnedt JT, Zhao Z, Schubert JR, Sen S. Insomnia symptoms and short sleep predict anxiety and worry in response to stress exposure: a prospective cohort study of medical interns. *Sleep Med* 2019;55:40-47.
- [50] Kassambara A. Practical Guide to Principal Component Methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra. : CreateSpace Independent Publishing Platform, 2017.
- [51] Kestler HA, Lausser L, Lindner W, Palm G. On the fusion of threshold classifiers for categorization and dimensionality reduction. *Comput Stat* 2011;26:321-340.
- [52] Klepstad P, Loge JH, Borchgrevink PC, Mendoza TR, Cleeland CS, Kaasa S. The Norwegian brief pain inventory questionnaire: translation and validation in cancer pain patients. *J Pain Symptom Manage* 2002;24:517-525.
- [53] Knaster P, Karlsson H, Estlander AM, Kalso E. Psychiatric disorders as assessed with SCID in chronic pain patients: the anxiety disorders precede the onset of pain. *Gen Hosp Psychiatry* 2012;34:46-52.

- [54] Kruskal W, Wallis W. Use of Ranks in One-Criterion Variance Analysis. *J Am Stat Assoc* 1952;47:583-624.
- [55] Kuhn M. caret: Classification and Regression Training; 2018.
- [56] Kurtze N, Rangul V, Hustvedt BE, Flanders WD. Reliability and validity of self-reported physical activity in the Nord-Trondelag Health Study: HUNT 1. *Scand J Public Health* 2008;36:52-61.
- [57] Landmark T, Dale O, Romundstad P, Woodhouse A, Kaasa S, Borchgrevink PC. Development and course of chronic pain over 4 years in the general population: The HUNT pain study. *Eur J Pain* 2018;22:1606-1616.
- [58] Lê S, Josse J, Husson F. FactoMineR: A R Package for Multivariate Analysis. *Journal of Statistical Software* 2008;25:1-18.
- [59] Leeuw M, Goossens ME, Linton SJ, Crombez G, Boersma K, Vlaeyen JW. The fear-avoidance model of musculoskeletal pain: current state of scientific evidence. *J Behav Med* 2007;30:77-94.
- [60] Lerman SF, Finan PH, Smith MT, Haythornthwaite JA. Psychological interventions that target sleep reduce pain catastrophizing in knee osteoarthritis. *Pain* 2017;158:2189-2195.
- [61] Liaw A, Wiener M. Classification and Regression by randomForest. *R News* 2002;2:18-22.
- [62] Lotsch J, Utsch A. Machine learning in pain research. *Pain* 2018;159:623-630.
- [63] MacQueen J. Some methods for classification and analysis of multivariate observations. Berkeley, California: University of California Press, 1967.
- [64] Maechler M, Rousseeuw P, Struyf A, Hubert M. Cluster: Cluster Analysis Basics and Extensions. 2017.

- [65] Mantyselka PT, Turunen JH, Ahonen RS, Kumpusalo EA. Chronic pain and poor self-rated health. JAMA 2003;290:2435-2442.
- [66] Markkula RA, Kalso EA, Kaprio JA. Predictors of fibromyalgia: a population-based twin cohort study. BMC Musculoskelet Disord 2016;17:29-6.
- [67] Mathias JL, Cant ML, Burke ALJ. Sleep disturbances and sleep disorders in adults living with chronic pain: a meta-analysis. Sleep Med 2018;52:198-210.
- [68] McCracken LM, Dhingra L. A short version of the Pain Anxiety Symptoms Scale (PASS-20): preliminary development and validity. Pain Res Manag 2002;7:45-50.
- [69] McCracken LM, Vowles KE, Eccleston C. Acceptance of chronic pain: component analysis and a revised assessment method. Pain 2004;107:159-166.
- [70] Milborrow S. rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'; 2018.
- [71] Molnar C. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Morrisville, NC, USA: Lulu Press, 2019.
- [72] Nedergaard M. Neuroscience. Garbage truck of the brain. Science 2013;340:1529-1530.
- [73] Newell A, Simon H. Computer science as empirical inquiry: symbols and search. Commun ACM 1976;19:113-126.
- [74] Pareto V. Manuale di Economia Politica. Milan: Società editrice libraria, 1909.
- [75] Partinen M, Gislason T. Basic Nordic Sleep Questionnaire (BNSQ): a quantitated measure of subjective sleep complaints. J Sleep Res 1995;4:150-155.

- [76] Pearson K. On lines and planes of closest fit to systems of point in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 1901;2:559-572.
- [77] Peltonen M, Harald K, Männistö S, Saarikoski L, Peltomäki P, Lund L, Sundvall J, Juolevi A, Laatikainen T, Aldén-Nieminen H, Luoto R, Jousilahti P, Salomaa V, Taimi M, Vartiainen E. Kansallinen FINRISKI 2007 -terveystutkimus: tutkimuksen toteutus ja tulokset. National Public Health Institute, Finland 2008.
- [78] Phillips N. yarrrr: A companion to the e-book: YaRrr!: The Pirate's Guide to R". 2017.
- [79] Pirkola S, Isometsa E, Aro H, Kestila L, Hamalainen J, Veijola J, Kiviruusu O, Lonnqvist J. Childhood adversities as risk factors for adult mental disorders: results from the Health 2000 study. Soc Psychiatry Psychiatr Epidemiol 2005;40:769-777.
- [80] R Development Core Team. R: A language and Environment for Statistical Computing; 2008.
- [81] Rand W. Objective Criteria for the Evaluation of Clustering Methods. Journal of the American Statistical Association 1971;66:846-850.
- [82] Reyes-Gibby CC, Aday L, Cleeland C. Impact of pain on self-rated health in the community-dwelling older adults. Pain 2002;95:75-82.
- [83] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77-77.
- [84] Roelofs J, McCracken L, Peters ML, Crombez G, van Breukelen G, Vlaeyen JW. Psychometric evaluation of the Pain Anxiety Symptoms Scale (PASS) in chronic pain patients. J Behav Med 2004;27:167-183.

- [85] Rousseeuw P. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Comp Appl Math* 1987;20:53-65.
- [86] Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics* 2007;23:2507-2517.
- [87] Saunders JB, Aasland OG, Babor TF, de la Fuente, J R, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. *Addiction* 1993;88:791-804.
- [88] Smith MY, Egert J, Winkel G, Jacobson J. The impact of PTSD on pain experience in persons with HIV/AIDS. *Pain* 2002;98:9-17.
- [89] Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Sattar A, Kang B. *Advances in Artificial Intelligence*. AI 2006. Lecture Notes in Computer Science, vol 4304. Berlin, Heidelberg: Springer; 2006:1015-1021.
- [90] Sørensen T. A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons. København: Kongelige Danske videnskabernes selskab, 1948.
- [91] Spearman C. The proof and measurement of association between two things. *The American Journal of Psychology* 1904;15:72-101.
- [92] Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* 2003;43:1947-1958.

- [93] Therneau T, Atkinson E. *rpart: Recursive Partitioning and Regression Trees*; 2019.
- [94] Turk DC, Rudy TE. Toward an empirically derived taxonomy of chronic pain patients: integration of psychological assessment data. *J Consult Clin Psychol* 1988;56:233-238.
- [95] Turk DC, Rudy TE. The robustness of an empirically derived taxonomy of chronic pain patients. *Pain* 1990;43:27-35.
- [96] Ultsch A, Lotsch J. Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data. *PLoS One* 2015;10:e0129767.
- [97] Vartiainen P, Heiskanen T, Sintonen H, Roine RP, Kalso E. Health-related quality of life and burden of disease in chronic pain measured with the 15D instrument. *Pain* 2016;157:2269-2276.
- [98] Vartiainen P, Mantyselka P, Heiskanen T, Hagelberg N, Mustola S, Forssell H, Kautiainen H, Kalso E. Validation of EQ-5D and 15D in the assessment of health-related quality of life in chronic pain. *Pain* 2017;158:1577-1585.
- [99] Vavrek M. fossil: palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica* 2011;14.
- [100] Venables WN, Ripley BD. *Modern Applied Statistics with S*. New York: Springer-Verlag; 2002.
- [101] Vincent A, Hoskin TL, Whipple MO, Clauw DJ, Barton DL, Benzo RP, Williams DA. OMERACT-based fibromyalgia symptom subgroups: an exploratory cluster analysis. *Arthritis Res Ther* 2014;16:463-7.
- [102] Ward J, Jr. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 1963;58:236-244.

- [103] Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2009.
- [104] Wicksell RK, Olsson GL, Melin L. The Chronic Pain Acceptance Questionnaire (CPAQ)-further validation including a confirmatory factor analysis and a comparison with the Tampa Scale of Kinesiophobia. *Eur J Pain* 2009;13:760-768.
- [105] Widerstrom-Noga EG, Duncan R, Turk DC. Psychosocial profiles of people with pain associated with spinal cord injury: identification and comparison with other chronic pain syndromes. *Clin J Pain* 2004;20:261-271.
- [106] Xie L, Kang H, Xu Q, Chen MJ, Liao Y, Thiyagarajan M, O'Donnell J, Christensen DJ, Nicholson C, Iliff JJ, Takano T, Deane R, Nedergaard M. Sleep drives metabolite clearance from the adult brain. *Science* 2013;342:373-377.
- [107] Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32-35.

Figure 1: Flow chart of data analysis. The figure gives an overview of the applied scientific approach to the data. After case selection (framed by a green dashed line), data analysis was performed in two main steps, which (i) aimed at the creation and interpretation of a pain-related phenotype cluster structure ("pain phenotype creation" and "pain phenotype interpretation"; framed by a blue dashed line), which was associated with non-pain-specific parameters in a second main step of the analysis (non-pain-specific "phenotype signal detection"; framed by a red dashed line), both of which included separate feature selection steps. In the flowchart, open rectangles represent cases (patients, case numbers coded with "n") or features (parameters, numbers coded with "d") or headings, while filled rectangles indicate the bioinformatic or statistical methods used. The first major analytical step was the **creation of clusters** in the data structure given by the five pain-related parameters, followed by **cluster interpretation** based on the pain-related characteristics used for cluster identification. Since pain-related characteristics were considered phenotype-defining variables, missing values were not imputed, but the corresponding patients were excluded. Clustering was performed using hierarchical clustering, followed by consolidation of the clusters using k-means clustering. Cluster interpretation was performed with algorithms of explainable artificial intelligence (XAI), which were implemented as classification and regression trees (CART) and partial decision trees (PART). For comparison, a standard non-XAI type classifier implemented as random forests was used. Feature selection, i.e. the selection of the most informative parameters, was implemented as a computed ABC analysis, which is an item categorization technique. CART-based hierarchical rules of cluster interpretation were finally created from the relevant features. The second main analytical step involved the **phenotype signal detection** among the parameters not directly related to pain. There, < 20% of missing data were imputed. The workflow of cluster assignment and feature selection was analogous to the second part of the first

analytical step. The figure has been created using Microsoft PowerPoint for Windows (version 2019, Microsoft Corporation, Redmond, WA, USA).

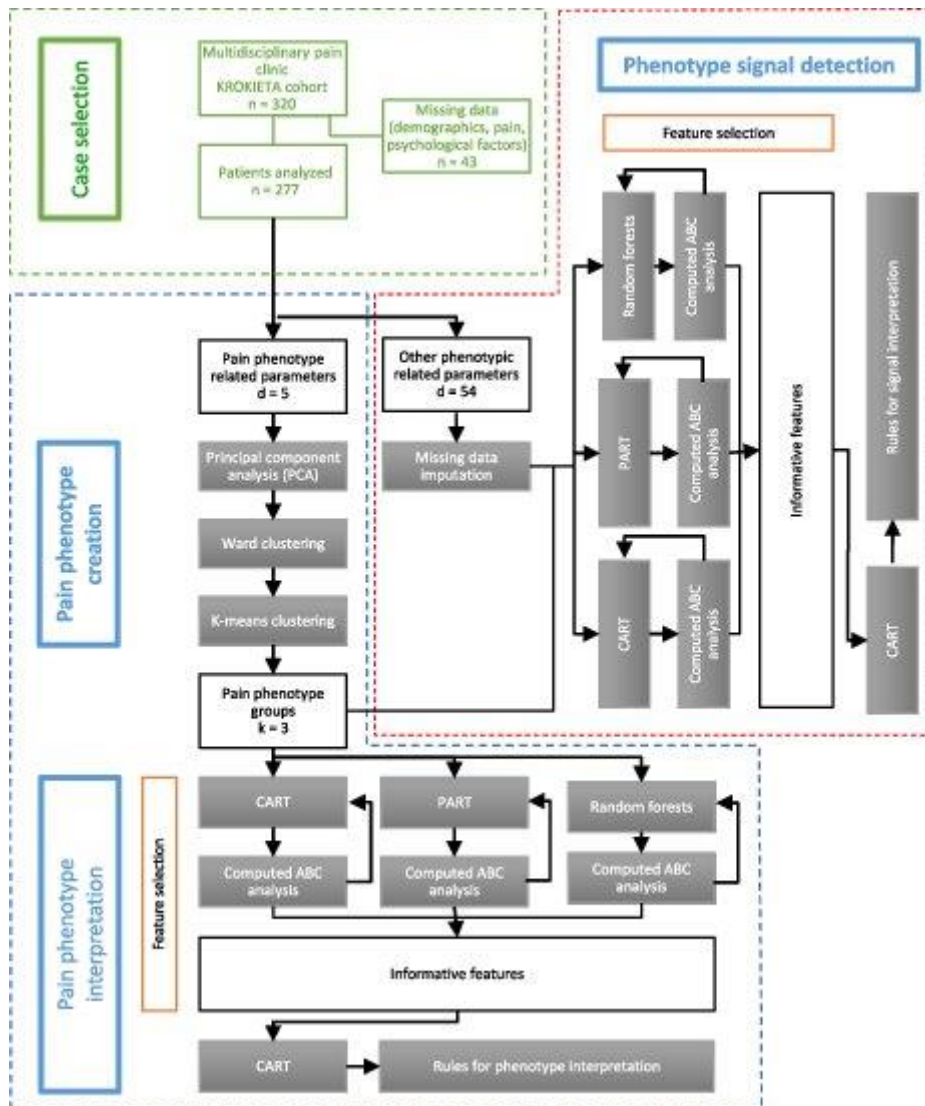


Figure 2: Clustering of the d = 5 pain phenotype-related parameters. A: Factorial plot of the individual data points on the principal component map, obtained following hierarchical Ward clustering and subsequent k-means based cluster consolidation. The colored areas visualize the cluster separation. **B:** Tree-shaped decision rules for the assignment of patients to the pain phenotype clusters based on the d = 2 parameters that prevailed as informative following feature selection performed during

interpretation of the clusters. The leaves at the bottom are color-coded for clusters. The first number in each leaf displays the number of correctly assigned cases with respect to cluster membership, while the second number displays the total number of cases assigned to this leaf. **C:** Original data of the pain phenotype-related parameters, shown as bean plots, separately for the three clusters. The individual observations are shown as black circles in a one-dimensional scatter plot, surrounded by the probability density function (pdf) of the distributions (colored areas). Box and whisker plots of the same data are overlaid on the bean plots. They have been constructed using the minimum, quartiles, median (solid black red line within the box), and maximum. The whiskers add 1.5 times the interquartile range (IQR) to the 75th percentile or subtract 1.5 times the IQR from the 25th percentile. The parameters found to be relevant for the pain-phenotype clustering are emphasized with red frames. The figure has been created using the R software package (version 3.6.1 for Linux; <http://CRAN.R-project.org/> [80]) and the R packages “yarr” (<https://cran.r-project.org/package=yarr> [78]) and “rpart.plot” (<https://cran.r-project.org/package=rpart.plot> [70] and “ggplot2” (<https://cran.r-project.org/package=ggplot2> [103])).

Pain intensity = mean score of Brief Pain Inventory pain intensity items (worst, least, average, right now). Affective pain interference = mean score of Brief Pain Inventory affective pain interference items (mood, relations with other people, enjoyment of life, sleep). Activity pain interference = mean score of Brief Pain Inventory activity pain interference items (walking, work, general activity).

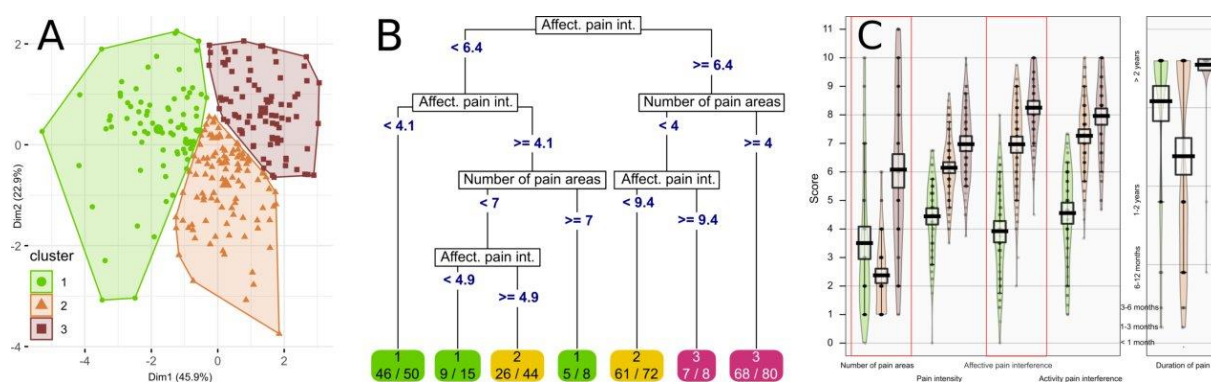


Figure 3: Tree-shaped decision rules for the assignment of patients to pain phenotype clusters #1, with lowest pain intensity and pain interference, (A), or #3, with high pain intensity and pain interference, and the greatest number of pain areas, (B), based on the phenotypic parameters from the six other categories not directly related to pain that had passed the feature selection step during tree building. The leaves at the bottom are color-coded for clusters. The first number in each leaf displays the number of correctly assigned cases with respect to cluster membership, while the second number displays the total number of cases assigned to this leaf. At the right of the tree, the original data with respect to the phenotypic parameters used for the decisions to assign a patient to the respective phenotype cluster are shown as bean plots. Data are shown separately for the cluster #1 versus the other clusters (top) or cluster #3 versus the other clusters (bottom). The individual observations are shown as black circles in a one-dimensional scatter plot, surrounded by the probability density function (pdf) of the distributions (colored areas). Box and whisker plots of the same data are overlaid on the bean plots. They have been constructed using the minimum, quartiles, median (solid black red line within the box), and maximum. The whiskers add 1.5 times the interquartile range (IQR) to the 75th percentile or subtract 1.5 times the IQR from the 25th percentile. The figure has been created using the R software package (version 3.6.1 for Linux; <http://CRAN.R-project.org/> [80]) and the R packages “rpart.plot” (<https://cran.r-project.org/package=rpart.plot> [70]) and “yarr” (<https://cran.r-project.org/package=yarr> [78]).

Syst. BP = Systolic blood pressure

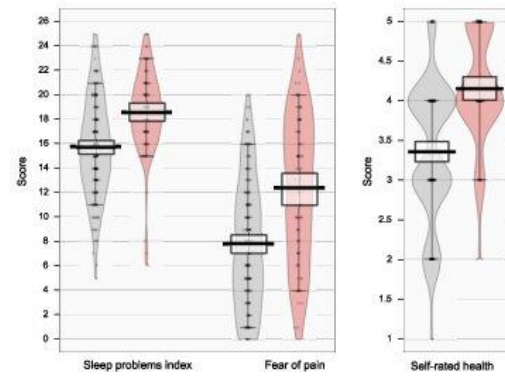
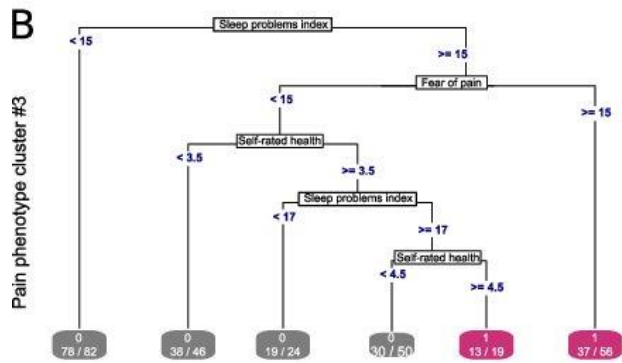
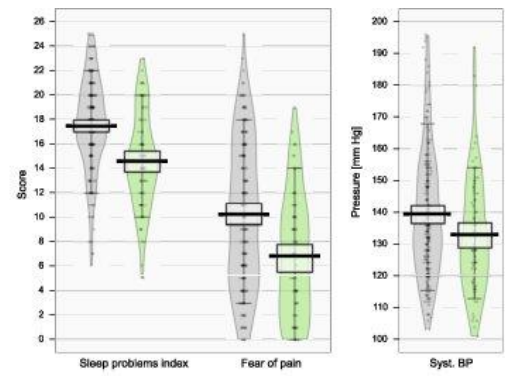
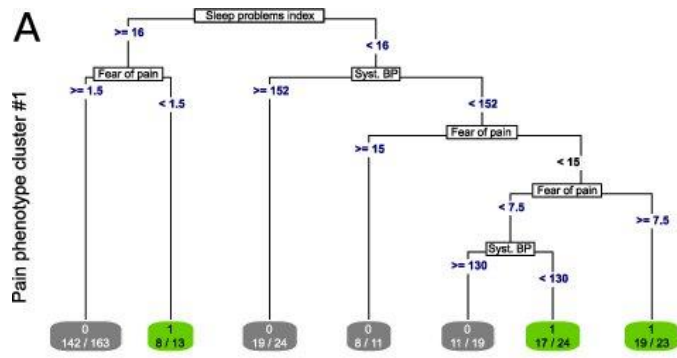


Table 1: Parameters acquired for the present analysis and basic descriptive statistics.

Parameter	N	Counts or descriptive statistical parameters	
Pain phenotype-related		Median	Interquartile range
Number of pain areas	277	3	2 - 5
Pain intensity	277	6	5 - 7
Affective pain interference	277	7	5 - 8.25
Activity pain interference	277	6.85	5.67 - 8.33
	N	Count	Value -
Duration of pain	277	1: 0 2: 4 3: 8 4: 26 5: 39 6: 200	1: < 1 month 2: 1-3 months 3: 3-6 months 4: 6-12 months 5: 1-2 years 6: > 2 years
Pain etiology-related		No	Yes
Any neuropathic pain	270	173	97
Low back pain	270	188	82
Musculoskeletal pain other than back pain	270	199	71
Facial pain	270	258	12
Abdominal pain	270	260	10
Complex Regional Pain Syndrome	270	254	16
Headache	270	263	7
Phantom limb pain	270	269	1
Fibromyalgia	270	241	29
Chronic pain syndrome	270	261	9
Other pain diagnosis	270	243	27
Psychological parameters		Median	Interquartile range
Self-rated health	277	4	3 - 4
BDI-II	277	14	9 - 21
Childhood adversities	277	2	1 - 3
CPAQ Activity engagement	277	33	24 - 42
CPAQ Pain willingness	277	17	11 - 22
PASS Cognitive anxiety	277	15	10 - 19
PASS Escape/Avoidance	277	13	9 - 16
PASS Fear of pain	277	9	4 - 14
PASS Physiological anxiety	277	8	4 - 12
Demographic parameters		Median	Interquartile range
Age in years	277	46	37 - 55
Number of children	277	2	0 - 2
Education in years	271	13	11 - 15
Household income	264	4	3 - 7
		Count	Code
Sex	277	96 men, 181 women	1: male 2: female
Civil status	276	1: 109 2: 1 3: 61 4: 67 5: 32 6: 6	1: married 2: registered relationship 3: cohabiting 4: unmarried 5: separated 6: widow
Type of work	277	1: 2 2: 19 3: 128 4: 20 5: 3	1: agriculture 2: manual work 3: office work 4: studying or at school 5: housewife

		6: 64 7:41	6: pensioner 7: unemployed
Disability pension	273	1: 191 2: 18 3: 32 4: 32	1: no 2: yes, part-time pension 3: yes, temporary pension 4: yes, permanent pension
Lifestyle-related		Median	Interquartile range
Exercise periods of >20 min per week	274	1	0 - 3
Hours spent sitting per day	265	6	3.5 - 9
Sleep problems index	273	17	14 - 20
Nutritional index	185	1	1 - 2
Body Mass Index	274	27.4	24.1 - 32.4
Systolic blood pressure, mmHg	275	134	124 - 150
Diastolic blood pressure, mmHg	275	86	77 - 94
Waist circumference	274	94.25	83.25 – 105.87
		Count	Code
Smoking currently	276	0: 165 1: 111	0: no 1: yes
Drug abuse	185	0: 168 1: 16 2: 1	0: no 1: has used 2: dependent
Alcohol consumption frequency	175	1: 36 2: 59 3: 55 4: 21 5: 4	1: never 2: once a month or less 3: 2-4 times a month 4: 2-3 times a week 5: 4 times a week or more
Previous treatments		Median	Interquartile range
Negative treatment experiences	277	3	1 - 4
Positive treatment experiences	277	4	2 - 6
Physician visits within previous 12 months	259	10	5 - 14
Missed work days within previous 12 months	247	35	0 - 180
Comorbidities		No	Yes
Hypertension	275	193	82
Heart failure	275	267	8
Angina pectoris	275	258	17
Diabetes	274	248	26
Asthma	275	230	45
Chronic obstructive pulmonary disease	275	267	8
Rheumatoid arthritis	275	271	4
Joint disease other than rheumatoid arthritis	275	199	76
Low back pain	275	128	147
Depression	273	197	76
Psychiatric disorder other than depression	275	254	21
Hypercholesterolemia ever in life	233	128	105
Using cholesterol medication	235	202	33
High blood pressure ever in life	274	156	118
Blood pressure medication use ever in life	120	39	81
		Count	Code
Diabetes type	277	1: 234 2: 10 3: 7 4: 19 5: 3 6: 4	1: no 2: no, but elevated blood sugar level 3: yes, type 1 diabetes 4: yes, type 2 diabetes 5: yes, but I don't know the type 6: yes, diabetes during pregnancy

Pain intensity = mean score of Brief Pain Inventory pain intensity items (worst, least, average, right now)

Affective pain interference = mean score of Brief Pain Inventory affective pain interference items (mood, relations with other people, enjoyment of life, sleep)

Activity pain interference = mean score of Brief Pain Inventory activity pain interference items (walking, work, general activity)

BDI-II = Beck Depression Inventory II

CPAQ = Chronic Pain Acceptance Questionnaire

PASS = Pain Anxiety Symptoms Scale

Table 2: Results of a principal component analysis (PCA) of the d = 5 pain phenotype-related parameters. Five principal components (PCs) were obtained. Components with eigenvalues > 1 (bolded) were retained for further analyses.

Parameter	Principal component				
PCA summary	PC1	PC2	PC3	PC4	PC5
Eigenvalue	2.29	1.14	0.83	0.41	0.32
% of explained variance	45.89	22.85	16.61	8.23	6.42
Cumulative % of explained variance	45.89	68.74	85.35	93.58	100.00
Variable loadings					
Number of pain areas	0.15	-0.67	-0.72	0.04	0.08
Duration of pain	0.11	-0.71	0.68	-0.11	0.06
Pain intensity	0.56	0.04	0.10	0.78	-0.26
Affective pain interference	0.57	0.08	-0.05	-0.59	-0.56
Activity pain interference	0.57	0.19	0.02	-0.16	0.78

Pain intensity = mean score of Brief Pain Inventory pain intensity items (worst, least, average, right now)

Affective pain interference = mean score of Brief Pain Inventory affective pain interference items (mood, relations with other people, enjoyment of life, sleep)

Activity pain interference = mean score of Brief Pain Inventory activity pain interference items (walking, work, general activity)

Table 3: Classification performance for the assignment to a pain phenotype-related cluster, obtained when training different classifiers (classification and regression trees (CART), partial decision trees (PART), random forests (RF)), multinomial log-linear regression (lReg) with the phenotypic parameters from the first category directly related to pain. Results represent the medians and 95% confidence intervals of the performance measures obtained during 1000 runs using random splits of the data set into disjoint training (2/3 of the data set) and test (1/3) data subsets, across the three pain phenotype clusters. The classifiers were trained on the full feature set of $d = 5$ pain phenotype-related parameters and on reduced sets with $d = 2$ pain-related parameters obtained following a feature selection step of the data analysis. Classifiers were trained on the original data set and again on a randomly permuted training data, which served as a negative control to detect possible overfitting.

Performance parameter	Full feature set				Reduced feature set							
	Original data				Original data				Permuted data			
Algorithm	CART	PART	RF	lReg	CART	PART	RF	lReg	CART	PART	RF	lReg
Number of included parameters	5	5	5	5	2	2	2	2	2	2	2	2
Sensitivity, recall [%]	81.5 (59.3 - 96.3)	86.1 (69 - 96.6)	89.7 (75.9 - 100)	100 (88.9 - 100)	74.1 (48.3 - 93.1)	75.9 (48.1 - 93.1)	75.9 (58.3 - 89.7)	75.9 (58.6 - 89.7)	33.3 (0 - 80.6)	6.9 (0 - 100)	33.3 (7.4 - 69)	18.5 (0 - 100)
Specificity [%]	91.1 (73.2 - 100)	92.9 (82.1 - 98.5)	95.4 (83.9 - 100)	100 (94.6 - 100)	87.7 (66.1 - 98.4)	89.2 (67.9 - 98.5)	89.2 (71.4 - 96.9)	88.9 (75 - 96.8)	69.2 (28.6 - 98.4)	93.7 (0 - 100)	69.2 (44.6 - 87.3)	87.3 (0 - 100)
Positive predictive value, precision [%]	81.5 (64.9 - 100)	85.7 (72.1 - 96.7)	91.7 (78.6 - 100)	100 (90 - 100)	74.4 (58.5 - 92.9)	76 (59.5 - 94.1)	75.9 (62.7 - 92.6)	75.9 (63.3 - 91.3)	35.9 (0 - 77.8)	39.1 (0 - 100)	35.7 (10.3 - 56.5)	42.8 (0 - 100)
Negative predictive value [%]	90.8 (79.7 - 98.2)	92.8 (83.1 - 98.4)	95.4 (89.6 - 100)	100 (94.8 - 100)	87.2 (73.8 - 96.6)	87.9 (73.8 - 96.6)	88.1 (77.6 - 95.2)	88.7 (76.6 - 95.2)	68 (47.8 - 84.9)	70 (41.5 - 100)	68.4 (53.1 - 78.6)	70.7 (40.1 - 100)
F1 [%]	80.8 (66.7 - 90.9)	85.2 (73.9 - 93.3)	90.4 (82.1 - 96.4)	98.2 (92.3 - 100)	73.5 (58.1 - 86.8)	74.1 (57.7 - 88.1)	75.3 (63.8 - 87.7)	76 (63.8 - 88.1)	35.6 (4.5 - 65.5)	55.1 (5.9 - 67.3)	34 (9.1 - 58.8)	55.6 (4.7 - 74.5)

Balanced accuracy [%]	85.4 (73.8 - 94)	89 (79.5 - 95.8)	92.3 (85.6 - 97.6)	98.5 (93.7 - 100)	79.9 (67.7 - 90.7)	80.4 (68.6 - 91.7)	81.3 (71.2 - 91)	82.1 (71.2 - 91.7)	50.7 (31.6 - 72.4)	50 (34.5 - 69.2)	51.1 (36.8 - 65.8)	50 (31.8 - 77.2)
AUC-ROC [%]	89.4 (81.1 - 94.8)	90.3 (83.7 - 95.5)	98.7 (97 - 99.6)	99.9 (99.2 - 100)	85.9 (79 - 91.7)	87.3 (80.9 - 92.3)	90.3 (86.3 - 94.2)	92.6 (89 - 95.6)	50.5 (32.6 - 68.7)	50 (35.1 - 65.1)	51.1 (36.6 - 66.3)	50.9 (13.4 - 87.3)
Discriminant power*	2.18 (1.24 - 13.13)	2.47 (1.6 - 13.56)	3.06 (2.16 - 14.28)	14.6 (3.32 - 25.39)	1.74 (9.9 - 2.72)	1.79 (0.97 - 2.89)	1.75 (1.1 - 2.84)	1.79 (1 - 2.79)	0.03 (0 - 1.46)	0 (0 - 11.7)	0.05 (- 0-95 - 0.86)	-0.24 (n.d. - 12.3)
Youden's index [%]	70.8 (47.5 - 88)	78 (58.9 - 91.7)	84.6 (71.2 - 95.2)	96.9 (88 - 100)	60.9 (35.7 - 80.1)	60.7 (37.3 - 83.3)	62.4 (42.5 - 82)	64.2 (42.4 - 83.3)	1.2 (-36.9 - 43.2)	0 (-30.9 - 38.5)	2.18 (-27.8 - 32.9)	0 (-36.5 - 53.7)

*According to [89], an algorithm is a poor discriminant if $DP < 1$, limited if $DP < 2$, fair if $DP < 3$, and good in all other cases.

n.d. = not defined.

Table 4: Classification performance for the assignment to a pain phenotype-related cluster, assessed for the complete three-cluster solution, and for each pain cluster separately versus the other clusters. The classification performance was assessed when training three different classifiers (classification and regression trees (CART), partial decision trees (PART), random forests (RF)) with (i) the phenotypic parameters from the first category directly related to pain (upper part of the table) and again, (ii) with phenotypic parameters from the six other categories not directly related to pain (lower part of the table). Results represent the medians and 95% confidence intervals of the performance measures obtained during 1000 runs using random splits of the data set into disjoint training (2/3 of the data set) and test (1/3) data subsets, across the three pain phenotype clusters. The classifiers were trained on the full feature set of $d = 5$ pain phenotype-related parameters and on reduced sets with $d = 2$ pain-related parameters obtained following a feature selection step of the data analysis. Classifiers were trained on the original data set and again on a randomly permuted training data, which served as a negative control to detect possible overfitting.

Parameter	Algorithm/ data	All pain clusters	Cluster #1	Cluster #2	Cluster #3
Pain features					
Features in ABC set "A"	CART	2	2	2	2
	PART	2	2	2	2
	RF	2	2	2	2
Balanced accuracy [%]					
CART	Orig.	79.9 (67.7 - 90.7)	86.6 (77.8 - 93.7)	70.9 (62.6 - 77.9)	83.7 (73.7 - 91.5)
	Perm.	50.7 (31.6 - 72.4)	50 (35.4 - 76.6)	50 (35.4 - 65.6)	50 (36.5 - 67.8)
PART	Orig.	80.4 (68.6 - 91.7)	86.3 (77 - 92.1)	71.4 (63.5 - 79.2)	79.7 (70.6 - 87.5)
	Perm.	50 (34.5 - 69.2)	50 (46.9 - 57.4)	50 (39.1 - 64.4)	50 (42.8 - 57)
RF	Orig.	81.3 (71.2 - 91)	87.7 (80.9 - 96.2)	70.9 (64 - 78.4)	85.8 (78.6 - 92.4)
	Perm.	51.1 (36.8 - 65.8)	50.3 (38 - 65.8)	51.9 (40.5 - 63.6)	52.7 (41.5 - 64.9)
Other features					

Features in ABC set "A"	CART	5	3	5	3
	PART	11	13	13	13
	RF	13	15	15	13
Balanced accuracy [%]					
CART	Orig.	61.4 (47.6 - 71.9)	64.5 (53.6 - 74.6)	59 (48.5 - 68.7)	64.6 (54.9 - 73.9)
	Perm.	50 (35.4 - 64.9)	50 (39.5 - 61.7)	49.7 (38.9 - 60.4)	50 (35.4 - 64.9)
PART	Orig.	62.1 (50.8 - 72.7)	60.8 (50.2 - 70.4)	60.9 (51.3 - 69.2)	67 (57.3 - 76.5)
	Perm.	50 (37.9 - 63)	49 (39.2 - 62.2)	49.5 (38.8 - 61.8)	50.8 (38.6 - 63)
RF	Orig.	64.2 (50.2 - 75)	65.3 (57.3 - 72.8)	57.6 (49.7 - 65.8)	65 (56.8 - 73.8)
	Perm.	50 (38 - 63.2)	50 (45.2 - 55.9)	49.2 (41 - 58.9)	50.1 (43.8 - 59.8)

ABC = computed ABC analysis

CART = Classification and Regression Trees

PART = Partial Decision Trees

RF = Random forests

Orig. = original data

Perm. = permuted data